

Kutatási adatok és kezelésük az irodalomtudományban (Szövegkiadások, adatbázisok, korpuszok)

Maróthy Szilvia

Kutatási terv a Magyar Tudományos Akadémia Bolyai János Kutatási Ösztöndíj 2020-as pályázatához

Bevezetés

Kutatási eredmények a számítógépes bölcsészeti kutatások mintegy hét évtizede során nemcsak hagyományos módon, tanulmányszövegekben kerültek rögzítésre, kifejtésre, hanem különféle digitális adat- és parancssorok formájában is. Ezek a WWW megjelenésével zömmel webes alkalmazások segítségével kerültek publikálásra, mely a grafikus böngészők óta az adatok grafikus megjelenítését is magával vonta. (A pályázat összefoglalójában nem térek ki a nem internetes elektronikus kutatási adatokra, pl. CD-ROM-okra, ám a kutatás esettanulmányaiban igen.)

Mára alig találkozni olyan nyertes (akár hazai, akár nemzetközi) kutatási pályázattal, mely ne vállalna valamilyen digitalizálási feladatot, kutatási eredmények internetes közzétételét. Ezek a tartalmak azonban jobbra szigetszerűen léteznek a weben, az egyes kutatási projektek csak ritkán tudnak egymásról, így a már elvégzett munkáról is. A projektek mentén elkészült webes tartalmak hamar gazdátlanokká válnak a futamidő lejártával, s a technika gyors fejlődése miatt egyúttal pillanatok alatt elavulttá. Számos példa hozható arra, hogy részben vagy akár teljesen elvesznek ezek a tartalmak, s velük együtt a jelentős tudományos munkák, elért kutatási eredmények.

Fontos kiemelni, hogy a kutatási adatkezelés elválaszthatatlan magától a kutatástól. A tudományos eredmények akkor válnak akceptálhatóvá, s kerülhetnek be a tudományos diskurzusba, ha a kutató(i közösség) a tudományos publikáció normáit betartja. A számítógépes feldolgozással és elemzéssel pedig a publikáció normái megváltoztak. Ezzel a változással kell az irodalom kutatóinak is lépést tartania.

A tudományos közlemények keletkezése, publikálása és hozzáférhetővé tétele egyaránt a webes közeghez igazodott, s ez nagy számú és különböző típusú digitális fájlok kezelését igényli. A GLAM szektorban, kivált egyetemi és kutatóhelyi könyvtárakban egyre általánosabbá váló repozitórium építési gyakorlat részint ezt a problémát hivatott kezelni. A könyvtáros és bölcsészettudományos gyakorlat egyaránt nehézségekkel küszködik, miként illeszse a hagyományos publikációknál kialakított minőség- és hitelesség-ellenőrző rendszerébe a hordozóhoz nem kötött, internetes közleményeket, ill. általában a digitális objektumokat. Az EU, s így Magyarország által az utóbbi évtizedekben több szabályozással, ajánlással támogatott szabad hozzáférés (open access) biztosítása is jelentős infrastrukturális háttérrel igényel, mely a mai napig sok esetben nem áll rendelkezésre. Hazánkban is vannak törekvések, elsősorban a könyvtáros szféra részéről a kutatási adatkezelés területén, ám ezek jelenleg igen sporadikusak. (Pl. openscience.hu és HUNOR; Micsik és Gárdos 2014.)

Az irodalomtudomány kutatóinak elemi érdeke, hogy a fenti kérdésekben maguk nyilvánuljanak meg, a tudomány sajátságai, teoretikus alapjai mentén egyedi igényeiket, problémáikat megfogalmazzák. Ha ezekre a sajátságokra a kutató közösség nem hívja fel időben a figyelmet, úgy a digitalizáció, a megőrzés és a kutatásértékelés felelős intézményei döntenek helyettük, vagy

feledkeznek meg róluk. A pályázó a fent ismertetett problémakörnek elsősorban az irodalomtudományos (tehát nem könyvtártudományi vagy tudománymetriai) részével foglalkozik, a számítógépes eszközökkel megtámogatott irodalomtudományi kutatások módszertani kérdéseiből indul ki. A pályázó ugyanakkor munkájával tevékenyen hozzájárul ahhoz, hogy a kutatást végző és ellenőrző, ill. a kutatói infrastruktúrákat működtető intézmények hatékony együttműködésbe kezdjenek.

A kutatás a következő témakörökre fókuszál:

- Mi az adat, kutatási adat irodalomtudományi kontextusban?
- Hogyan keletkeznek és vesznek el irodalomtudományi adatok? (pl. szövegkiadások, adatbázisok)
- Mit jelent a kutatási adatmenedzsment (Research Data Management) és az adatkezelési terv (Data Management Plan)?
- Hogyan kell kutatási adatokat kezelni? (Szabványok, ajánlások és implementálhatóságuk irodalomtudományi kutatásokra.)
- Az irodalomtudományi kutatási adatok tudománymetriai, kutatásértékelésben elfoglalt helye.

A fenti témakörök kutatása, beható ismerete az elméleti kérdések körüljárásán, megoldásán túl megkövetelik egyes részterületek gyakorlati megvalósítását is. Ennek érdekében a pályázó esettanulmányokon keresztül mutatja be, az elméleti megfontolások a gyakorlatban miként érvényesíthetők. A kutatás ezen szakaszában a Szegedi Tudományegyetem egyes kutatócsoportjaival és az egyetem Klebelsberg Kuno könyvtárával, valamint az MTA Könyvtár és Információs Központtal kerül sor együttműködésre. A kutatás célkitűzése, hogy mintát szolgáltasson irodalomtudományi kutatási adatok (szövegkiadások, adatbázisok és adatsorok, szövegkorpuszok és elemzésükre szolgáló alkalmazások) kezelésére: szabványos kialakítására, repozitóriumi archiválására, dokumentálására, szabad hozzáférésű közzétételére, az újrafelhasználhatóság biztosítására.

Témakörök

Mi az adat, kutatási adat irodalomtudományi kontextusban?

Ahhoz, hogy kutatási adatok létrehozásáról, közléséről, archiválásáról beszéljünk, elengedhetetlen a *kutatási adat* fogalmának definiálása. Ez azonban nem magától értetődő – a kutatás során a következő kérdésekre kell választ találni: Csak a számítógépes eszközökkel rögzített adatsorokat, szövegeket, jegyzeteket értjük-e alatta, s ha igen, minden ilyen a szövegszerkesztőben írt jegyzetektől az SQL adattábláig? Csak a rendezett adatokat, vagy minden szöveges tartalmat, ami a kutatás során keletkezik? Milyen viszonyban áll a publikált kutatási eredmény és az annak alapjául szolgáló kutatási adat? Ha a kutatási adat nem magában, hanem webes alkalmazással kerül közzétételre, vagy a kutatás eredményének részét képezik a számítógépes forráskódok is, mit tekintünk archiválandónak, az adott formában rögzített adatsorokat, vagy az alkalmazás egészét?

A pályázó korábbi munkásságában a tudományos szövegkiadások kapcsán már foglalkozott a fenti kérdéskörrel, így a jelen kutatás során ezek eredményeire is épít. (Lásd pl. a szakmai életrajzban hivatkozott: „Elektronikus szövegkiadások a könyvtárban”, és „Tudományos szövegkiadások a hálózaton: Áttekintés” című cikkeket.) A pályázó jelenleg is aktívan segíti az MTA–BTK Lendület Magyar Irodalom Politikai Gazdaságtana Kutatócsoport tagjaként annak digitalizálási tevékenységét, így gyakorlati tapasztalatokkal is rendelkezik.

A kutatási adat definícióját a tudományos szövegkiadásokon túl az irodalomtudományi adatbázisokra, az adatközpontú irodalomtörténeti kutatásokra, a stilometriai és topic modelling elemzésekre tekintettel kell kialakítani.

Hogyan keletkeznek és vesznek el irodalomtudományi adatok?

A jellemzően projektek keretében elkészült internetes tudományos szövegkiadások, adatbázisok a futamidő lejártával gyakran magukra maradnak. Már kialakításukkor is ritkán törekednek a projektvezetők más webes platformokkal, tartalmakkal való összekapcsolódásra, ill. olyan szabványos és nyílt forráskódú környezet kialakítására, mely szavatolná a fenntarthatóságot, megőrzést. Emiatt mára számos webes kiadvány technikailag elavult, jelenleg nem használható, vagy el is vészett (példákat szolgáltat erre a pályázó fent idézett írásain túl „A tudományos elektronikus szövegkiadások feldolgozása”, és „A vonzó véletlen” című cikke).

2020-ban először a nemzetközi, ill. EU-s pályázatokhoz hasonlóan az NKFIH/OTKA pályázatoknál is követelmény a kutatási adatkezelési terv megléte. Ennek elméleti és gyakorlati alapjai azonban a hazai bölcsészettudományi kutatásokban hiányoznak, így a szabályozást betartók és betartatók egyaránt segítségre szorulnak.

Mit jelent a kutatási adatmenedzsment és az adatkezelési terv?

A kutatási adatmenedzsment vagy adatkezelés (Research Data Management) célja, hogy gondoskodjon a tudományos eredmények hozzáférhetővé tételének biztosításáról. Ez magában foglalja azok létrehozásakor a szabványok és ajánlások figyelembe vételét, a keletkezett kutatási adatok közreadását, ellenőrizhetővé, hivatkozhatóvá és újrafelhasználhatóvá tételét, valamint azok tartalmi és formai megőrzését a jövő nemzedékek számára. Ezek együttesen igen komplex elvárásrendszer képeznek, melynek természetes módon a kutatócsoport, vagy a kutatói környezetet biztosító intézmények egyedül nem, csak közösen tudnak eleget tenni. A kutatási adatok szabványos megosztását támogatandó az Európai Bizottság nyílt tudománnyal foglalkozó munkacsoportja 2016-ban közzétette a FAIR alapelveket (Wilkinson et al. 2016). A FAIR betűszó a Findable, Accessible, Interoperable és Reusable szavakból áll össze, az alapelveket ma már több nemzetközi (pl. ERC) kutatócsoport is igyekszik alkalmazni. Jelen pályázat keretében sor kerül a FAIR alapelvek áttekintésére, az irodalomtudományban való implementálására, ill. ennek módszertani kérdéseire. (A kutatási adatok közreadásával kapcsolatban vö. a pályázó „A nyílt és a zárt tudományról” című cikkét.)

A FAIR alapelvek ahhoz is segítséget nyújtanak, hogy az egyes kutatók, kutatócsoportok milyen szempontok szerint készítsék el adatkezelési tervüket (Data Management Plan). Az adatkezelési terv leírja az adott kutatás során gyűjtött, feldolgozott, ill. generált adatok kezelésével kapcsolatos tevékenységet, összefüggésben az adatkezelési szabályozással, ajánlásokkal – ez esetben a FAIR alapelvekkel. Egy kutatási terv részét képező adatkezelési tervnek a következőket kell tartalmaznia: adatok kezelése a projekt futamideje alatt és azt követően; milyen adatokat gyűjt, dolgoz fel, ill.

generál a kutató(csoport); milyen módszertanok és szabványok kerülnek alkalmazásra; az adatok szabad hozzáférés szerint kerülnek-e közzétételre; az adatokat miként gondozzák, őrzik meg. (Vö. European Commission, 2016.)

Hogyan kell kutatási adatokat kezelni?

A tudományos közösségek gyakorlata nemcsak a kutatási adat definíciójában, de annak kezelésében is eltérő. A pályázat célkitűzése, hogy áttekintse, milyen nemzetközi gyakorlatok vannak az irodalomtudomány, ill. a bölcsészettudományok területén – amilyen például a 2020 februárjában megjelent ALLEA-ajánlás (Harrower et al. 2020) –, s milyen módszertani megfontolások mentén érdemes egy kutatási projekt adatmenedzsmentjét megszervezni. Hazai (Lendület, OTKA) kutatócsoportokkal együttműködve ugyanakkor lehetővé válik a teoretikus megfontolások gyakorlatba ültetése is. A kutatás ezen részének középpontjában az említett FAIR alapelvek és irodalomtudományi alkalmazásuk áll.

A kutatási adatok megtalálható, hozzáférhető, együttműködésre és újrafelhasználásra alkalmas közzététele azonban igen komplex elvárásrendszer jelent a gyakorlatban. A FAIR alapelvek megvalósításáról beszámoló első közlemények, melyek az utóbbi 1-2 évben jelentek meg (pl. Tóth-Czifra, 2019), azt mutatják, az egyes elvárások olykor egymással ütközőek is lehetnek. Az elvárásrendszer összetettsége egyben a kutatók és a kutatói környezet, infrastruktúra számára is kihívásokat rejt, olyan ismeretek elsajátítását követeli meg, melyek a hagyományos publikációs közegben nem voltak szükségesek (pl. Open Access-konstrukciók, közzétételi licencek, időtálló számítógépes formátumok, ISO és más nemzetközi szabványok ismerete).

A futó kutatási projektek segítése mellett a pályázat keretében archiválásra kerül a *Balassi Bálint összes verse, hálózati kritikai kiadás* (Horváth-Tóth-Vadai, 1994–2016) összes ismert és még fellelhető verziója. Ez mintegy 5-6 különféle formátumú és kódolású (pl. HTML, TEI XML) verziót jelent, melyek az elmúlt közel két és fél évtizedben keletkeztek. A kiadások tudományos értéke, filológiai és történeti fontossága vitathatatlan, mégis csak töredékéhez fér hozzá a kutatói közösség, részint a készítő, részint a befogadó intézmények mulasztásai miatt. A kiadások összegyűjtése, archiválása és közzététele mellett szükséges azok minél pontosabb leírása. A pályázó korábbi kutatásai során ezen szövegkiadások történetét áttekintette, az egyes kiadványok összegyűjtését megkezdte, ám eddig az archiváláshoz szükséges infrastruktúra nem állt rendelkezésére, így ennek megvalósítására is a pályázat keretein belül kerülhet sor. (Lásd „A tudományos elektronikus szövegkiadások feldolgozása”, és „A vonzó véletlen” című cikkeket.)

Az irodalomtudományi kutatási adatok tudománymetriai, kutatásértékelésben elfoglalt helye

A hazai kutatásértékelés mérvadó adatbázisa, melyet alapvetően a kutatók teljesítményének mérhetővé tétele érdekében készítettek és tartanak fenn, a Magyar Tudományos Művek Tára (MTMT). Az MTMT igyekszik tekintettel lenni a tudományterületi specifikumokra, mégis egyes hiányosságai miatt a kutatói tevékenység számszerűsítése, osztályozása olykor nem áll összhangban a valós kutatói tevékenységgel, kivált a bölcsészettudományok területén.

A számítógépes eszközökkel rögzített, feldolgozott, ill. tárolt kutatási eredmények ugyan leírhatók az MTMT rendszerében, ám tudománymetriai mutatókban a jelentőségük elenyésző. Míg például a hagyományos könyv formában publikáló szerző jelentősen javítja mutatóit, addig egy kutatói adatbázist akár évtizedekig fejlesztő és fenntartó kutató alig profitál a tudományos értékét és

mennyiségét tekintve legkevésbé sem különböző eredményeiből. A közlemény és a közleményekre történő hivatkozások elszámolása tekintetében egyaránt hátrányban van ma az a magyarországi kutató, aki korszerű technikai eszközökkel végzi és teszi közzé kutatásait. A kutatás ezekre a problémákra is fel kívánja hívni a figyelmet, segítve ezzel a kutatói közösség mellett az MTMT-t fejlesztő, ill. a rendszert a kutatásértékelésben használó intézményeket.

Hasznosulás

Jelenleg Magyarországon a bölcsészettudományi kutatások számára szinte teljesen hiányoznak azok az infrastrukturális és intézményes feltételek, melyek alkalmazkodnak az elmúlt 20–30 évben lezajlott mediális változásokhoz a kutatási eredmények publikálásában. Ennek legfőbb oka, hogy az ezzel kapcsolatos elméleti reflexió is gyakorta hiányzik a bölcsészettudományi diskurzusból. A pályázat keretében ezen elmaradások pótlására kerülhet sor, elsősorban az irodalomtudomány területét érintő kérdésekben. A kutatás eredményei ugyanakkor nagy mértékben hasznosíthatók más bölcsészettudományi környezetben is, ami a számítógépes feldolgozás és elemzés kutatási eredményeit és azok közzétételét, megőrzését illeti.

A kutatási eredmények tudományos és technológiai feltételeknek egyaránt megfelelő közzététele a kutatói közösség vitathatatlan érdeke. A tudományos eredményekhez való hozzáférés ma már jelentős mértékben elektronikus, zömmel hálózati hozzáférést jelent. A kutatómunka részét kell képezze, hogy oly módon kerüljenek közlésre a kutatási eredmények, hogy a tudomány alap értékei az új mediális közegben ne sérüljenek. Ilyen alapvető érték a hitelesség, s az ezt szolgáló hivatkozhatóság, megismételhetőség, ellenőrizhetőség, s ilyen érték a tudás szabad áramlása. Ma ezek a körülmények sok esetben nem állnak fenn, ezen kíván a pályázó kutatásaival változtatni.

A számítógépes irodalomtudományi kutatások tekintetében a teoretikus, módszertani alapok is olykor hiányoznak. Az MTA Textológiai Munkabizottsága például a 2000-es évek eleje óta (Alapelvek, 2004) nem reagál a hagyományos filológiai módszertan mellett egyre nagyobb teret foglaló számítógépes filológiára, mely nem pusztán eszközbeli, hanem valóban paradigmatisztikus változást hozott magával. A kutatói közösségek figyelmét a pályázat keretében elért eredmények képesek felhívni arra, hogy ezzel a közel 30 éve zajló változással behatóbban foglalkozzanak. A kutatás eredményei konkrét fogódzókat, esettanulmányokat, módszertani javaslatokat nyújthat ehhez.

A kutatás jelentős mértékben hozzájárul a Magyar Tudományos Akadémia által is hangsúlyosan támogatott szabad hozzáférés (open access) technológiai és kutatómódszertani alapjainak megteremtéséhez, melyben a hazai bölcsészettudományi közösség elmaradásai mára nyilvánvalóak. (Vö. Az MTA Open Access-rendeletei, 2012 és 2016.)

Munkaterv

A 2020–2021-es év első felében a pályázó áttekinti a vonatkozó hazai és nemzetközi szakirodalmat, ezzel párhuzamosan megkezdheti a munkát az SZTE Klebelsberg Könyvtárával és az MTA Könyvtárral. A könyvtárakkal való együttműködés során közösen kialakítják azt a környezetet és munkamódszert, mellyel a kutatási adatkezelés a gyakorlatban megvalósíthatóvá válik.

A 2020–2021-es év második félévében a megismert nemzetközi gyakorlatok mentén kezdi meg a munkát az együttműködő kutatócsoportokkal, melyek a kutatás esettanulmányait adják. Emellett összegyűjti a Balassi-kiadás (Horváth–Tóth–Vadai, 1994–2016) egyes verzióinak számítógépes állományait, megkezdheti a könyvtárral együttműködésben annak feldolgozását. Konferencia-

előadásban és tanulmányban adja közre addigi kutatási eredményeit. Szakrecenziót és ismeretterjesztő írást közöl a kutatási témához kapcsolódóan. Eredményeit és gyakorlati tapasztalatait műhelyfoglalkozás (workshop) keretében adja át a kutatói közösség számára. Az SZTE Klebelsberg Könyvtár könyvtáros továbbképzésén előadást tart a könyvtáros közösséget érintő kérdésekben, pl. a könyvtári ökoszisztéma miként támogathatja a kutatási adatkezelést.

A 2021–2022-es év első felében a pályázó befejezi a Balassi-kiadások archiválási munkálatait.

Az együttműködő kutatócsoportoknak folyamatos segítséget nyújt módszertani kérdésekben, a kutatási adatok archiválását segíti. Az archiválási munkák tapasztalatait, módszertani javaslatait magyar és angol nyelvű szaktanulmányokban foglalja össze, magyar és angol nyelvű konferencia-előadások formájában bocsátja vitára. Szakrecenziót és ismeretterjesztő írást közöl a kutatási témához kapcsolódóan. Az SZTE Klebelsberg Könyvtár könyvtáros továbbképzésén előadást tart a könyvtáros közösséget érintő kérdésekben, pl. a kutatócsoportokkal való együttműködés tapasztalatait foglalva össze.

2022-ben a kutatás lezárásaként workshopot szervez a témában jártas további szakértők bevonásával. A műhelyfoglalkozás részint előadásokból, részint gyakorlati foglalkozásokból állna, melyek a számítógépes irodalomtudomány és a kutatási adatkezelés témakörét ölelik fel.

Vállalások: 2020–2021

- egy magyar nyelvű publikáció (téma: Kutatási adatmenedzsment az irodalomtudományban: alapfogalmak, módszertan)
- egy magyar nyelvű konferencia-előadás hazai tudományos konferencián
- egy recenzió a témában
- egy ismeretterjesztő cikk a témában
- egy workshop megtartása a témában (előadás és gyakorlat)
- egy előadás az SZTE Klebelsberg Könyvtár könyvtáros továbbképzésén

Vállalások: 2021–2022

- egy magyar nyelvű publikáció (téma: Kutatási beszámoló az áttekintett módszertani kérdésekről, esettanulmányokkal a gyakorlati megvalósításokról)
- egy idegen nyelvű publikáció (téma: FAIR alapelvek implementálása az irodalomtudományban: esettanulmányok és módszertani reflexiók)
- egy magyar nyelvű konferencia-előadás hazai tudományos konferencián
- egy idegen nyelvű konferencia-előadás nemzetközi tudományos konferencián
- egy recenzió a témában
- egy ismeretterjesztő cikk a témában
- egy workshop megtartása a témában (előadás és gyakorlat)
- egy előadás az SZTE Klebelsberg Könyvtár könyvtáros továbbképzésén

Hivatkozott irodalom

„A Magyar Tudományos Akadémia elnökének 24/2016. (VII. 15.) számú határozata: A tudományos művek nyílt hozzáférésű közzétételéről”, 2016.

http://openaccess.mtak.hu/dokumentumok/mta_határozat_oa_2016.pdf.

„A Magyar Tudományos Akadémia elnökének 27/2012. (IX. 24.) számú határozata: Tudományos művek nyílt hozzáférésű közzétételének irányelvei”, 2012.

https://mta.hu/data/dokumentumok/hatteranyagok/akademiai_szabalyozasok/27_2012_elnoki_hat_Open_Access5.pdf.

„Alapelvek az irodalmi szövegek tudományos kiadásához”. *Irodalomtörténet* 35, sz. 3. sz. (2004): 328–30.

European Commission, Directorate-General for Research & Innovation. „Guidelines on FAIR Data Management in Horizon 2020”, 2016.

https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

Harrower, Natalie, Maciej Maryl, Timea Biro, és Beat Immenhauser. *Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities*. Berlin: ALLEA – All European Academies, 2020. <https://doi.org/10.7486/DRI.tq582c863>.

Horváth Iván, Tóth Tünde, és Vadai István, szerk. *Balassi Bálint összes verse, hálózati kritikai kiadás, v1.0 és v2.0*. Hálózati kritikai kiadássorozat. Gépeskönyv, 1998. <http://magyar-irodalom.elte.hu/gepesk/bbom/>.

Horváth Iván, Tóth Tünde, és Vadai István, szerk. *Balassi Bálint összes verse, hálózati kritikai kiadás, v3.0*. Gépeskönyv, 2004. online megjelenítés elveszett.

Horváth Iván, Tóth Tünde, Vadai István, és Parádi Andrea, szerk. *Balassi Bálint összes verse, hálózati kritikai kiadás, v4.0*. TEI XML, 2016. publikálatlan.

HUNOR (HUNGarian Open Repositories), <https://openscience.hu/hunor/>.

Maróthy Szilvia. „Elektronikus szövegkiadások a könyvtárban”. *Tudományos és Műszaki Tájékoztatás* 64, sz. 6 (2017): 298–309.

Maróthy Szilvia. „Tudományos szövegkiadások a hálózaton: Áttekintés”. *Irodalomtörténeti Közlemények* 122, sz. 5 (2018): 617–633.

Maróthy Szilvia. „A tudományos elektronikus szövegkiadások feldolgozása”. In *Valóságos könyvtár – könyvtári valóság, III*, szerkesztette Kiszl Péter és Csík Tibor, 351–356. Budapest: ELTE BTK Könyvtár- és Információtudományi Intézet, 2018.

Maróthy Szilvia. „A vonzó véletlen”. In *HI70/tanítványok: Tanulmányok Horváth Iván 70. születésnapjára*, szerkesztette Bartók Zsófia Ágnes, Bognár Péter, és Maróthy Szilvia. Q. E. D. Kiadó, 2018.

Maróthy Szilvia. „A nyílt és a zárt tudományról”. In *Kulturális iparágak, kánonok és filterbuborékok*, szerkesztette Bárány Tibor et al., Budapest: BME–ELTE, 2020.

Micsik András, és Gárdos Judit. „Tudományos repozitóriumok az MTA-ban: a KDK és a SZTAKI tanulságai”. In *Informatika a felsőoktatásban*. Debreceni Egyetem Informatikai Kar, 2014. <http://openarchive.tk.mta.hu/340/>.

Open Science, Debreceni Egyetem, <https://openscience.hu/>.

Tóth-Czifra, Erzsébet. „10. The Risk of Losing the Thick Description: Data Management Challenges Faced by the Arts and Humanities in the Evolving FAIR Data Ecosystem”. In *Digital Technology and the Practices of Humanities Research*, szerkesztette Jennifer Edmond, 235–66. Open Book Publishers, 2020. <https://doi.org/10.11647/obp.0192.10>.

Wilkinson, Mark D., et al. „The FAIR Guiding Principles for Scientific Data Management and Stewardship”. *Scientific Data* 3 (2016. március 15.): 160018. <https://doi.org/10.1038/sdata.2016.18>.